DOCUMENT RESUME

ED 191 885 TM 800 509

AUTHOR Tobin, Kenneth G.; Capie, William
TITLE The Development and Validation of a Group Test of Logical Thinking.
PUB DATE Apr 80
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Tests; *Group Testing; Higher Education; Intermediate Grades; *Logical Thinking; *Problem Solving; Secondary Education; *Test Construction; Test Reliability; *Test Validity; Videotape Recordings
IDENTIFIERS Inhelder Piaget System; *Test of Logical Thinking (Tobin and Capie)

ABSTRACT

The development of the Test of Logical Thinking (TOLT), a group test of formal reasoning ability that requires students to solve problems and to justify the solutions obtained, is presented. Procedures are described for investigations of internal consistency, construct validity, and criterion-related validity. Items used in prior research serve as a basis for the development of TOLT, assuring that the test contains valid measures of formal reasoning ability. Two of the items measure each of five modes of formal reasoning: (1) controlling variables; (2) proportional reasoning; (3) probabilistic reasoning; (4) correlational reasoning; and (5) combinatorial reasoning. High school students' unsuccessful formulation of clearly written justification for selection of particular responses led to modification of the initial version of TOLT; thus providing multiple justifications as well as multiple solutions. The revised test uses a color videotape to present the context for each problem and to standardize administrative procedures. Evidence suggests that TOLT does measure formal thought. Reliability data indicate high internal consistency, and validity data support an effective group test of formal thought. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED191885

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE DEVELOPMENT AND VALIDATION OF A GROUP TEST OF LOGICAL THINKING

Kenneth G. Tobin

and

William Capie

University of Georgia
Department of Science Education
Athens, Georgia 30602

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. Tobin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the
American Educational Research Association Meeting

April, 1980

Boston

TM 800509

THE DEVELOPMENT AND VALIDATION OF A GROUP TEST OF LOGICAL THINKING

INTRODUCTION

Initially studied by Piaget and his colleagues, the development of formal reasoning ability has been extensively researched in adolescents and adults (e.g., Lovell, 1961; Farrell, 1969; Arlin, 1975; Chiappetta, 1976). In the majority of cases, clinical interviews based on protocols described by Inhelder and Piaget (1958, 1975) have been used to assess formal reasoning ability. Two important trends that have emerged from research are that many adolescents and adults are limited in their ability to use formal modes of reasoning and formal reasoning ability is an important mediator of cognitive achievement (e.g., Cantu and Herron, 1978; Goodstein and Howe, 1978). As a consequence, researchers have emphasized the importance of modifying instructional objectives, materials, and activities so that they are suited to the cognitive development of learners. Concomitantly they have urged that priority be given to the development of formal reasoning ability of middle and high school students through the use of appropriate curriculum materials.

Action on either of these concerns requires that formal reasoning ability be assessed in a valid and reliable manner. The clinical interview procedure is not suited for widespread administration because of the time required to administer a set of tasks to a group of students and because of the level of expertise required of the interviewer. In a research context an additional problem derives from the subjectivity of the procedures used. The advantages of simultaneously testing large groups of subjects have led to a number of attempts to develop valid

group measures of formal reasoning ability (e.g., Burney, 1974; Tisher and Dale, 1975; Lawson, 1978; Staver and Gabel, 1979). Items from most of these tests were based on content and logical processes derived from the works of Inhelder and Piaget (1958, 1975).

Ease of administration and objective scoring procedures are inducements for researchers to develop a valid and reliable pencil and paper measure of formal reasoning ability. A major difficulty, however, is to ensure that subjects use formal reasoning ability to solve the items on the test. Whether subjects use formal reasoning or not may be ascertained from their reasons for developing or choosing a response. A characteristic of the clinical procedures that has generally not been incorporated into pencil and paper tests is the necessity for subjects to justify their solution to a problem. Exceptions to this were the test items described by Lawson (1978) and Lawson, Adi and Karplus (1979), which required subjects to provide written justifications for their solutions to problems. Another important aspect of the procedure used by Lawson was the use of demonstrations to provide a realistic context for the problems to be solved.

The purpose of this study was to develop a group test of formal reasoning ability that would require students to solve problems and justify the solutions obtained. Development of the Test of Logical Thinking (TOLT) commenced with a selection of ten items previously reported by Lawson (1978) and Lawson et al (1979).

PROCEDURES

In the sections that follow, procedures are described for the development of TOLT, and investigations of internal consistency, construct validity and criterion-related validity.

Development of TOLT

Items that had been used in prior research (Lawson, 1978; Lawson et al, 1979) were used as a basis for developing an initial version of TOLT. The adoption of this procedure assured that TOLT would contain items that had been previously reported as valid measures of formal reasoning ability. Two items were selected to measure each of five modes of formal reasoning: controlling variables, proportional reasoning, probabilistic reasoning, correlational reasoning and combinatorial reasoning. The test incorporated demonstrations to provide a context for the items. Students selected a correct response from a number of alternatives and provided written justification for their selection.

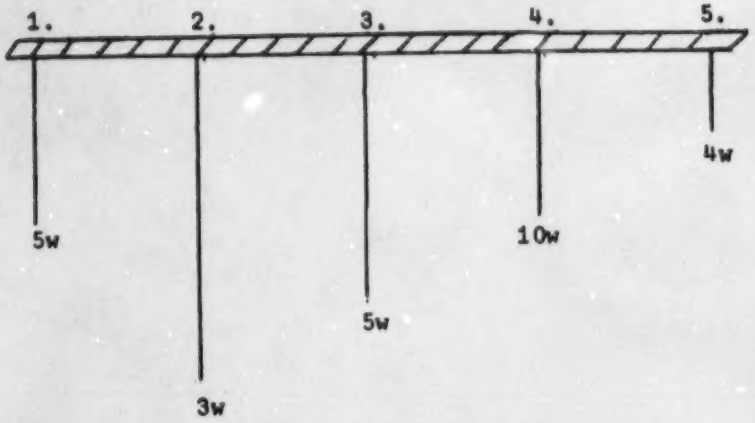
Although the reliability of TOLT was reasonably high (α = .74) it was apparent that several factors were reducing validity. For example, many high school students were unable to formulate a clear written justification for selecting a particular response. Rescoring these responses suggested many inconsistencies in scoring. Because of these problems, the initial version of TOLT was modified so that multiple justifications were provided as well as multiple solutions for each problem. Reasons that had been volunteered by subjects on the first version of TOLT were modified and incorporated into a revised version of the test. In the revised test a correct solution required selection of the correct response and the best justification for the response. A sample item is included in Figure 1 to illustrate the item format.

(Include Figure 1 About Here)

The revised test also utilized a color video-tape to present the context for each problem and to standardize administration procedures.

Item 4

The Pendulum's Weight



Suppose you wanted to do an experiment to find out if changing the weight on the end of the string changed the amount of time the pendulum takes to swing back and forth. Which pendulums would you use for the experiment?

a. 1 and 4

b. 2 and 4

c. 1 and 3

d. 2 and 5

e. all

Reason

1. The heaviest weight should be compared to the lightest weight.
2. All pendulums need to be tested against one another.
3. As the number of washers is increased the pendulum should be shortened.
4. The number of washers should be different but the pendulums should be the same length.
5. The number of washers should be the same but the pendulums should be different lengths.

Figure 1. A sample item from TOLT

Adequate time was provided for each item to be completed. The time required for the test to be administered was approximately 40 minutes.

Investigations of TOLT Validity and Reliability

Three samples were used in an investigation of the reliability and validity of TOLT. TOLT was administered to a sample of 353 students in middle school grades 6, 7 and 8; a sample of 82 physics and chemistry students from grades 11 and 12; and, 247 students enrolled in college science courses.

Intact classes were used in all cases to obtain data for reliability and construct validity investigations. In such studies the purpose is not to differentiate subjects at different grade levels, but to identify the extent to which test items measure a common dimension. An adequate investigation of these test characteristics requires a range in ability on the construct being measured. As a consequence, subjects were selected from a number of grade levels.

Prior research (Renner and Grant, 1978) had shown the developmental levels of physics students to be distributed differently than developmental levels of all students. A higher proportion of physics students were found to be operating at a formal operational level. As a consequence, students from grade 11 and 12 physics and chemistry classes were selected in an endeavor to obtain a greater range in formal reasoning. It was not intended that the sample be representative of the population of grade 11 and 12 students.

The internal consistency of TOLT was assessed from the complete data set ($n = 682$) using coefficient α (Cronbach, 1951).

Two separate factor analyses were conducted in an investigation of the underlying structure of the performance data. In the first, the data to be analyzed consisted of performance on each of the five modes of formal reasoning. In the second analysis, performance on each item was analyzed.

In each factor analysis a principal axis procedure was used to extract common factors. In this procedure, diagonal elements of the correlation matrix of the five modes of reasoning were replaced with communality estimates. The initial estimates were the squared multiple correlation of each mode of reasoning with the remaining modes of reasoning. A scree test (Cattell, 1966) and the Kaiser criterion (Kaiser, 1960) were jointly applied to obtain a final solution to the factor analysis. Rotation for interpretation was not required as one-factor solutions were obtained in each case.

Criterion-Related Validity of TOLT. The sample for this investigation consisted of 25 students enrolled in college science education courses. TOLT was administered to students followed by a battery of five clinical interviews selected from those described by Inhelder and Piaget (1958, 1975) to provide a measure on each of the formal modes of reasoning assessed by TOLT. On the basis of performance on each task students were rated one if they demonstrated formal thought in solving a problem or zero if they did not. Ratings for each task were summed to provide a measure of performance on the clinical interviews. Descriptions of the protocols used are provided in the Appendix.

Subject performance on the interviews was correlated with performance on TOLT. In addition, the relationships between performance on the interviews and TOLT was separately investigated for each mode of reasoning. The correlation coefficients obtained from these analyses provided a measure of the criterion-related validity of TOLT. The magnitude of the true relationship between the measures was investigated by applying a correction for the attenuation that occurred because of the unreliability of TOLT measures.

Performance on the interviews was also used as a means of identifying scores on TOLT that could be used to categorize subjects as being at the formal, transitional or concrete stage of development. Subjects were categorized as formal on the interview tasks if formal reasoning ability was used to correctly solve three of the five tasks. If subjects used formal reasoning on only one or two of the tasks they were categorized transitional. TOLT scores were selected such that misclassification was minimized.

RESULTS

Coefficient α for TOLT, based on the total sample of 682 students was .85. The internal consistency of each two-item subtest ranged from .56 to .82. Descriptive data related to the items and subtests of TOLT are contained in Table 1. Item difficulties ranged from .18 to .41 with an average of .30. Item discrimination indices ranged from .39 to .71 with an average of .55.

(Include Table 1 About Here)

When the data were separately analyzed for each grade level a gradual increase was evident from grade 6 to college level. A frequency distribution for each grade level in the sample is provided in Table 2.

(Include Table 2 About Here)

Performance on the five modes of reasoning was moderately inter-correlated. The correlation coefficients included in Table 3 demonstrate the inter-related nature of performance on the five modes of formal reasoning incorporated into TOLT.

(Include Table 3 About Here)

The factor analysis of the measures on each of the five modes of reasoning produced a one-factor solution which accounted for 43 percent of the common variance. Each mode of reasoning was highly correlated with the factor, with the factor structure loadings ranging from .60 to .72. The factor structure loadings and communality estimates for this analysis are contained in Table 4.

(Include Table 4 About Here)

In the second factor analysis a one-factor solution which accounted for 38 percent of the common variance was obtained. In this case the factor structure loadings ranged from .49 to .73 (Table 5) and were indicative of a common structure underlying performance on each item.

TABLE 1

Descriptive Data for TOLT

	Possible Score	Mean	Standard Deviation	Coefficient α
Item 1: Proportional Reasoning	1	.24	.42	
Item 2: Proportional Reasoning	1	.18	.38	
Subtest: Proportional Reasoning	2	.41	.75	.82
Item 3: Controlling Variables	1	.33	.47	
Item 4: Controlling Variables	1	.36	.48	
Subtest: Controlling Variables	2	.69	.87	.82
Item 5: Probabilistic Reasoning	1	.33	.47	
Item 6: Probabilistic Reasoning	1	.38	.48	
Subtest: Probabilistic Reasoning	2	.71	.81	.61
Item 7: Correlational Reasoning	1	.41	.49	
Item 8: Correlational Reasoning	1	.28	.45	
Subtest: Correlational Reasoning	2	.69	.78	.56
Item 9: Combinatorial Reasoning	1	.21	.41	
Item 10: Combinatorial Reasoning	1	.23	.42	
Subtest: Combinatorial Reasoning	2	.44	.73	.71
TOTAL	10	2.94	2.94	.85

TABLE 2

Distribution of Formal Reasoning Ability

Number Correct on TOLT	Grade 6 (%)	Grade 7 (%)	Grade 8 (%)	Grade 11 (%)	College (%)
0	47	43	34	11	9
1	25	29	23	9	6
2	20	15	23	5	15
3	5	8	7	7	15
4	2	2	4	6	11
5	1	0	5	11	7
6	0	1	2	2	10
7	0	1	1	9	9
8	0	0	1	17	8
9	0	0	1	7	7
10	0	0	0	16	4
Sample Size	100	86	167	82	247
Mean	0.9	1.1	1.6	5.5	4.4
Standard Deviation	1.1	1.4	1.8	3.4	2.8

TABLE 3

Inter-correlations among Modes of
Formal Reasoning on TOLT

	1	2	3	4	5
1. Proportional reasoning	1.00				
2. Controlling variables	.49	1.00			
3. Probabilistic reasoning	.50	.43	1.00		
4. Correlational reasoning	.42	.41	.41	1.00	
5. Combinatorial reasoning	.53	.35	.46	.42	1.00

TABLE 4

Factor Structure Loadings for each
Mode of Formal Reasoning

	Factor Structure Loading	Communality
Proportional reasoning	.72	.52
Controlling variables	.62	.38
Probabilistic reasoning	.66	.44
Correlational reasoning	.60	.36
Combinatorial reasoning	.66	.43

(Include Table 5 About Here)

Criterion-Related Validity

The correlation between performance on the interviews and TOLT was .76. Because of a restricted range of formal reasoning ability in the college sample, it was not possible to obtain a dependable estimate of the reliability of the clinical interviews. As a consequence, correction for attenuation was only applied for unreliability in TOLT. A validity coefficient of .82 was obtained. This index is indicative of a strong relationship between performance on the two measures. Descriptive data for each measure are contained in Table 6.

(Include Table 6 About Here)

Misclassification was minimized when TOLT scores of four and one were used as a basis for classifying subjects as formal and transitional respectively. A TOLT score of one was regarded as insufficient evidence of a developing formal reasoning capability because of the possibility of guessing a correct solution.

Eighty-eight percent of the subjects were classified at the same level of cognitive development using TOLT as they were with the clinical interviews. The categorical data are contained in Table 7.

(Include Table 7 About Here)

PREDICTIVE VALIDITY

Results from concurrent investigations at the University of Georgia (Table 8) indicated significant relationships between TOLT performance and performance on other variables. These studies provide an indication of the predictive validity of TOLT.

TABLE 5

Factor Structure Loadings
for TOLT Items

	Factor Structure Loading	Communality
Item 1: Proportional Reasoning	.72	.52
Item 2: Proportional Reasoning	.73	.53
Item 3: Controlling Variables	.66	.44
Item 4: Controlling Variables	.64	.40
Item 5: Probabilistic Reasoning	.63	.40
Item 6: Probabilistic Reasoning	.50	.25
Item 7: Correlational Reasoning	.49	.24
Item 8: Correlational Reasoning	.52	.27
Item 9: Combinatorial Reasoning	.63	.39
Item 10: Combinatorial Reasoning	.58	.33

TABLE 6

Descriptive Data for Performance on Interviews and TOLT

	Interview		TOLT		
	Mean	Standard dev.	Mean	Standard Dev.	Correlation
Proportional reasoning	.20	0.41	1.04	0.89	.29
Controlling variables	.96	0.20	1.80	0.50	.33*
Combinatorial reasoning	.60	0.50	1.00	0.91	.37*
Probabilistic reasoning	.56	0.51	0.80	0.51	.34*
Correlational reasoning	.68	0.48	1.04	0.98	.52**
TOTAL	3.00	1.19	5.68	2.64	.76***

* $p < .05$

** $p < .01$

*** $p < .001$

TABLE 7

Categorization Based on Performance
on TOLT and Clinical Interviews

		Interview		
		Concrete	Transitional	Formal
TOLT	Concrete		1	
	Transitional		6	1
	Formal		1	16

(Include Table 8 About Here)

In a study involving 353 students from grades 6 through 8 (Tobin and Capie, 1980) thirty-five percent of the variance on a test of integrated science processes was attributable to variation in performance on TOLT. Attenuation correction suggested a validity coefficient of .74.

Tobin, Capie and Bradley (1980) found that the performance of 150 college students was significantly correlated with integrated science process achievement ($r = .49$, $p < .0001$).

Yeany, Helseth and Barstow (1980) reported significant relationships between TOLT performance, achievement in college biology and scores obtained on the Scholastic Aptitude Test (SAT).

DISCUSSION

TOLT provides a reliable means of assessing formal reasoning ability. The ten item test has a high internal consistency and several of the subtests are sufficiently reliable to allow decision making at the subtest level. The reliability coefficients are of sufficient magnitude to enable the test to be used diagnostically, or in a research context, in studies designed to promote specific formal reasoning abilities. The magnitude of intercorrelations among modes of formal reasoning were suggestive of a common underlying structure. When the coefficients were corrected for the attenuation that occurs because of the unreliability of the subscales, the correlations ranged from .46 to .70.

Each of the modes of formal reasoning was highly correlated with the one-factor solution obtained from the factor analysis of subtest scores. This result can be interpreted in terms of each mode of reasoning contributing to a common underlying structure.

TABLE 8

Predictive Validity Data for TOLT

Author(s)		Dependent Measure	Grade Level	Correlation	Probability
Tobin and Capie	1979	Integrated Process Skills	8 Pretest	.57	.0001
		Integrated Process Skills	6 - 8 Summative	.59	.0001
		Integrated Process Skills	6 - 8 Retention	.57	.0001
Tobin, Capie and Bradley	1980	Integrated Process Skills	College Pretest	.49	.0001
		SAT (Quantitative)	College	.64	.0001
		SAT (Verbal)	College	.44	.0001
Tobin, Riley and Capie	1980	Integrated Process Skills	College Pretest	.41	.026
		Integrated Process Skills	College Posttest	.59	.003
Yeany, Helseth and Barstow	1980	Human Biology	College	.54	.0001
		Genetics	College	.55	.0001
		SAT (Quantitative)	College	.66	.0001
		SAT (Verbal)	College	.33	.005

A similar interpretation is applicable to the factor analysis of the separate item data. These results provide support for the construct validity of TOLT. The data suggest that each item provides a measure on an underlying dimension which is defined to be formal reasoning ability.

Although a comparatively small sample was involved in the criterion-related validation of TOLT, a high non-chance relationship was established between TOLT performance and clinical interview performance. A correlation of .82 between the two variables is indicative of similar performance on each variable. The evidence suggests that the same processes that allow subjects to correctly solve problems in the clinical interview are involved in the solution of TOLT items.

The use of TOLT performance as a means of categorizing students as formal, transitional or concrete provides a means of comparing results obtained using TOLT with studies conducted using clinical interviews or other instruments. The 88 percent agreement obtained between categorization using TOLT performance and categorization using interview performance provides evidence of the validity of the procedure. The 12 percent misclassification represents three subjects misplaced by one cell in the classification matrix. This estimate may be somewhat conservative based on the transitional characteristics of the subjects.

The data obtained in the investigation of criterion-related validity suggest that TOLT is measuring formal reasoning ability. Additional studies of this type are required using larger samples which encompass the entire range of formal reasoning ability.

The initial purpose of developing TOLT was to construct an instrument that could be used in studies of teaching and learning. On the basis of psychometric properties described in the paper, the use of TOLT is

advocated to obtain measures that differentiate student formal reasoning ability. Data so obtained would be useful in relating formal reasoning ability to achievement, investigating possible interactions of formal reasoning ability with teacher variables, or statistically controlling for variations in formal reasoning so that the effects of other teacher and student variables can be determined. In these respects TOLT can be used to obtain continuous or categorical measures of formal reasoning ability. From the data presented in Table 8 it is evident that TOLT has already been used for the purposes specified above.

While the development of TOLT grew from practical concerns relating to research and diagnostic teaching, the methodology used to develop TOLT is applicable to further investigate the nature of formal reasoning. Specific tests for each mode of reasoning could be developed, or existing TOLT items could be modified to enable additional information to be obtained from each item. For example, a set of follow-up questions could be constructed. The principle of using multiple multiple-choice formats offers the potential for developing measures of formal reasoning that are valid and reliable. With these instruments based on a sound clinical interview experience, researchers would have efficient tools to study developmental patterns among large numbers of subjects.

CONCLUSION

Evidence suggests that TOLT does measure formal thought. The reliability data are indicative of high internal consistency and the validity data are diverse and supportive of an effective group test of formal thought.

TOLT provides a means of assessing formal reasoning ability as a

diagnostic aid for teachers or as data for researchers investigating the nature of learning. The test is suitable for administration on a group basis to students from grade 6 through college.

REFERENCES

Arlin, P. Cognitive development in adulthood: a fifth stage. Developmental Psychology, 1975, 11, 602-606.

Burney, G. M. The construction and validation of an objective formal reasoning instrument. Unpublished doctoral dissertation, University of Northern Colorado, 1974.

Cantu, L. & Herron, D. Concrete and formal Piagetian stages and science concept attainment. Journal of Research in Science Teaching, 1978, 15, 135-143.

Cattell, R. B. The Scree test for the number of factors. Multivariate Behavioral Research, 1966, 1, 245-76.

Chiappetta, E. A review of Piagetian studies relevant to science instruction at the secondary and college level. Science Education, 1976, 60, 253-261.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Farrell, M. A. The formal stage: a review of the research. Journal of Research and Development in Education, 1969, 3, 111-118.

Goodstein, M. & Howe, A. The use of concrete methods in secondary chemistry instruction. Journal of Research in Science Teaching, 1978, 15, 361-366.

Inhelder, B. & Piaget, J. The growth of logical thinking from childhood to adolescence. New York: Basic, 1958.

Inhelder, B. & Piaget, J. The origin of the idea of chance in children. New York: N. W. Norton and Company, 1975.

Kaiser, H. F. The application of electronic computers to factor analysis. Educational and Psychological Measurement, 1960, 20, 141-151.

Lawson, A. E. The development and validation of a classroom test of formal reasoning. Journal of Research in Science Teaching, 1978, 15, 11-24.

Lawson, A. E., Adi, H. & Karplus, R. Development of correlational reasoning in secondary schools: do biology courses make a difference? The American Biology Teacher, 1979, 41, 420-425.

Linn, M. C. & Pulos, S. Predicting formal operational reasoning. Paper presented at the American Educational Research Association Meeting, April 1979, San Francisco, California.

REFERENCES (Cont.)

Lovell, K. A follow-up study of Inhelder and Piaget's the growth of logical thinking. British Journal of Psychology, 1961, 52, 142-153.

Renner, J. & Grant, R. Can students grasp physics concepts? The Science Teacher, 1975, October, 30-33.

Staver, J. R. & Gabel, D. L. The development and construct validation of a group-administered test of formal thought. Journal of Research in Science Teaching, 1979, 16, 535-544.

Tisher, R. P. & Dale, L. G. Understanding in science test. Victoria: Australian Council for Educational Research, 1975.

Tobin, K. G. and Capie, W. The effects of student engagement rate, formal reasoning ability and locus of control in middle school science achievement. Paper presented at the Annual Meeting of the Eastern Educational Research Association, Norfolk, Virginia, March, 1980.

Tobin, K. G., Capie, W. & Bradley, C. The relationship formal reasoning ability to process skill learning with college students. Paper in preparation, 1980.

Tobin, K. G., Riley, J. & Capie, W. An investigation of the relationships between teachers' use of formal operations and science process skill acquisition. Paper in preparation, 1980.

Yeany, R., Helseth, E. H. & Barstow, W. Interactive instructional video-tapes, scholastic aptitude, cognitive development and locus of control as variables influencing science achievement. Paper to be presented at the 1980 meeting of the National Association for Research in Science Teaching.

APPENDIX

The protocols used in the assessment of formal reasoning ability using the clinical interview procedure were selected from tasks described by Inhelder and Piaget (1958, 1975). One task was selected for each mode of reasoning pertinent to the development of TOLT. A description of the administration procedures for each task is provided below.

Equilibrium in the Balance

This task assessed ability to utilize proportional reasoning. The subject was presented with a 50g mass suspended four units from the fulcrum of a balance and asked to achieve a balance using a 50g mass, a 20g mass and a 10g mass, If the initial problems were successfully solved the subject was asked to use more than one mass to balance the beam on which a 20g mass was positioned seven units from the fulcrum. The final task required the subject to balance a 50g mass placed seven units from the fulcrum. In this task, formal thinking is characterized by an ability to reason that an increase in mass can be compensated by a proportional decrease in the distance from the fulcrum.

Pendulums

This task assessed the ability to separate variables and independently assess the effect of each on the period of a pendulum. The subject was presented with a set of three pendulums to which weights could be added or removed, and the string lengths altered. Subjects were initially asked to identify the factors likely to determine the time taken for one complete swing of the pendulum. Adopting a procedure used by Linn and

Pulos (1979) subjects were asked questions that required them to control, criticize, analyze and name variables. The four question types measured different aspects of the formal ability of keeping all other things equal while investigating the effects of a variable. To be assessed as formal, a subject needed to exhibit behaviors which satisfied each question type. Examples of the types of questions that were used for this task are provided in Figure 2.

(Include Figure 2 About Here)

Combinatorial Reasoning

This task assessed the ability to construct a full set of combinations using sequences of colored disks. In the first problem students were provided with three colored disks and asked to state the number of different ways they could be arranged. If a correct response was obtained, a fourth disk was added and subjects were again asked to nominate the number of possible different arrangements. After subjects responded to the question they were asked to explain how they arrived at the specified number, If necessary they were able to manipulate the disks or use pencil and paper to assist in solving the problem. To be assessed as formal the subject needed to systematically or mathematically obtain the correct answer and provide a satisfactory explanation of the solution obtained.

Probabilistic Reasoning

The objects used in this task were white disks, some of which had a red star on one side. Two sets of disks were formed and the subject was asked to identify the set from which there was a higher probability of selecting a disk with a star on it. Ten categories of questions were used to assess this mode of reasoning.

Name the variables that affect the period of a pendulum.

How would you assess the effects of different lengths on the period of a pendulum?

How would you assess the effects of different weights on the period of a pendulum?

How would you assess the effects of different angles of release on the period of a pendulum?

Could this set-up be used to test the effect of length on the period of a pendulum?

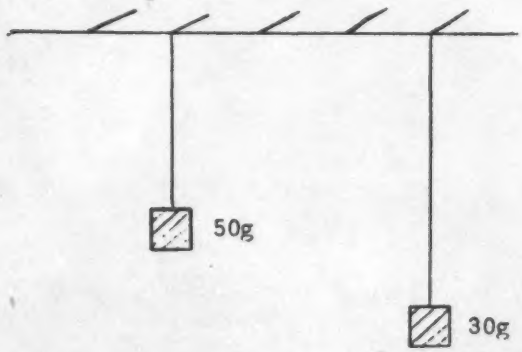


Figure 2. Examples of questions used in the pendulums task

For each question the composition of the two sets to be compared was varied. Examples of the ten categories are provided in Figure 3. Not all of the first six categories were asked, however questions from categories seven through 10 were asked of all subjects who had successfully answered questions from prior categories. To be categorized as formal, subjects had to solve problems for each category that was presented. Subjects were given multiple opportunities to demonstrate the required modes of thinking.

(Include Figure 3 About Here)

Correlational Reasoning

Four sets of cards which differed in their combination of hair and eye coloration were used to assess formal reasoning in this task. Six questions were formulated and presented to assess correlational thinking. The question types, with examples of the referent picture card sets are contained in Figure 4.

(Include Figure 4 About Here)

Category	Group 1 Stars	Composition Total Disks	Group 2 Stars	Composition Total Disks
Double impossibility	0	2	0	3
Double certainty	2	2	3	3
Certainty - impossibility	2	2	0	2
Possibility - certainty	1	2	1	1
Possibility - impossibility	1	2	0	2
Identical composition	1	3	1	3
Proportionality	1	2	2	4
Inequality of favorable cases/ equality of possible cases	1	4	2	4
Equality of favorable cases and inequality of possible cases	1 2	2 4	1 2	3 5
Inequalities respectively of favorable cases and possible cases without proportionality	1 2	2 3	2 6	3 7

Figure 3. Examples of problems used to assess probabilistic reasoning

Question	Blue eyes Blond hair	Blue eyes Brown hair	Brown eyes Blond hair	Brown eyes Brown hair
Is there a relationship between eye color and hair color?	5	1	2	4
Is there a relationship between eye color and hair color?	6	0	0	6
Show me a group where there is no relationship?	4 3	2 3	2 1	4 5
In which of these groups is the relationship between eye color and hair color stronger?	3	3	1	5
How would you alter group composition to make the relationship stronger?				
How did you decide which card to remove?				
What is the best way to decide if there is a relationship between hair color and eye color?				

Figure 4. Examples of questions used to assess correlational reasoning